{...}

The Monte Carlo fallacy

— Or, why all articles in the *New England Journal of Medicine* whose conclusions are announced on NPR are essentially bullshit: in all statistical studies of relatively small populations which attempt to establish correlations between "causes" (e.g. variations in diet) and "effects (e.g. cancer of the colon), a fallacy is involved more or less equivalent to the one that would be present in an attempt to, e.g., measure the value of pi by Monte Carlo methods by placing m points at random in a unit cube of n dimensions and trying to count the number within the unit sphere[1] — but where m and n are of the same order of magnitude. That is, the number of dimensions of the relevant phase space are at least as many as the data points in these "studies"; and I don't care how may bogus safeguards they think they're building

[1] Actually throwing darts at an n-sphere is not a good example, since for dimension greater than 10 the relative measure of the sphere decreases rapidly, and the Monte Carlo estimate of pi is usually zero. But the situation with an arbitrary function of hundreds of variables could only be worse. — One might note, e.g., that polynomials of degree 4 with 14 variables can represent arbitrary recursive sets.

into the analysis, the method obviously doesn't work, or they wouldn't keep reversing their findings every few years.[2]

---

[2] The post-genomic version of the fallacy has already been exploded: once it became easy to analyze genomes it was discovered that, contrary to expectation, most genetic disorders depend on contributions from hundreds or thousands of genes, and the idea that some single controller can be isolated and adjusted by the action of a single drug is almost always wishful thinking. — One example was the premature identification [K.P. Lesch, D. Bengel, A. Heils, S.Z. Sabol., B.D. Greenberg, S. Petri, J. Benjamin, C.R. Müller, D.H. Hamer, and D.L. Murphy, "Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region": *Science*. **274** (5292), 1527–31 (November 1996)] of variations in the gene SLC6A4, which codes for a serontonin transporter, as the trigger for depression and anxiety disorders; after a literature of several hundred papers had grown up around the conjecture, a systematic analysis of purported correlations of this and 17 other candidate genes with much larger statistical samples [Richard Border, Emma C. Johnson, Luke M. Evans, Andrew Smolen, Noah Berley, Patrick F. Sullivan, and Matthew C. Keller, "No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples"; *AJP in Advance* (doi: 10.1176/appi.ajp.2018.18070881)] revealed the conclusion to have been premature. — The latter authors conclude "the genetic underpinnings of common complex traits such as depression appear to be far more complicated than originally thought," and note that similar studies have refuted candidate gene hypotheses for schizophrenia and white matter microstructure.